Inferring transition rates on networks with incomplete knowledge

Purushottam D. Dixit*
Department of Systems Biology, Columbia University, New York, NY 10032

Abhinav Jain and Gerhard Stock
Institute of Physics and Freiburg Institute for Advanced Studies (FRIAS) Albert Ludwigs University, Freiburg, 79104 Germany

Ken A. Dill
Laufer Center for Physical and Quantitative Biology, Department of Chemistry, and Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY, 11790

Across many fields, a problem of interest is to predict the transition rates between nodes of a network, given limited stationary state and dynamical information. We give a solution using the principle of Maximum Caliber. We find the transition rate matrix by maximizing the path entropy of a random walker on the network constrained to reproducing a stationary distribution and a few dynamical averages. A main finding here is that when constrained only by the mean jump rate, the rate matrix is given by a square-root dependence of the rate, $\omega_{ab} \propto \sqrt{p_b/p_a}$, on p_a and p_b, the stationary state populations at nodes a and b. We give two examples of our approach. First, we show that this method correctly predicts the correlated rates in a biochemical network of two genes, where we know the exact results from prior simulation. Second, we show that it correctly predicts rates of peptide conformational transitions, when compared to molecular dynamics simulations. This method can be used to infer large numbers of rates on known networks where smaller numbers of steady-state node populations are known.

INTRODUCTION

We are interested in an inference problem in network science. Given the topology of a network and stationary populations at the nodes, what is the best model that can infer the rates of the dynamical flows along the edges? Here are examples. First, consider a spin model with a known stationary distribution, for example, those used in neuroscience (1), protein evolution (2), or colloidal sciences (3). It is of great interest to infer the best dynamical process that is consistent with a given rate of spin flip. Second, in systems biology, we often know the topology of a network of metabolites, or proteins or regulatory elements. In addition, "-omics" experiments can estimate the abundances of the many metabolites or proteins or regulatory elements at the nodes during the steady-state functioning of a cell. However, this information alone is not sufficient to explain cell function. We also need to know the forward and backward rates of fluxes ω_{ab} and ω_{ba} between all nodes a and b, for example in metabolic networks (4). Measuring all these rates is practically impossible at present, particularly for large networks. Third, in structural biology, it is common to perform computer simulations of the conformations of biomolecules and infer Markov models among metastable states from those simulations (5). Here, computing the populations of the states can be done rapidly, whereas computing the kinetic barriers between them is much slower.

In many such problems, the popular approach, especially for large networks, is to hypothesize a parametric dynamical model and learn the parameters of this model from data. Often, a large amount of data is required and the parameters learnt are not unique. We treat this problem as a matter of inference, in the spirit of statistical mechanics, where full distribution functions are inferred from a few measured equilibrium-averaged properties (6, 7). We provide a solution employing the dynamical analog of the principle of Maximum Entropy, Maximum Caliber, seeking the single best model that is consistent with under-determined data.

Mathematically, we seek a Markovian random walker on a network that has the maximal path entropy and that otherwise satisfies prescribed constraints. Towards that goal, we first define a class of walkers that satisfy a) a prescribed stationary state distribution $\{p_a\}$ over the nodes $\{a\}$ of a network and b) certain dynamical properties defined over the ensemble of stationary state paths $\{\Gamma\}$. Examples of dynamical properties include the average distance travelled by the walker per unit time, the average number of reactions per per unit time, or the average number of amino acid changes per unit time in

a constantly evolving protein.

Below, we first derive the Maximum Caliber Markov process. We then illustrate its predictive power with two examples, a gene expression network and a network of metastable states of a small peptide.

THEORY

Consider a Markovian random walker on a directed network G with nodes $V = \{a\}$ and edges E. Assume that the random walker has a unique stationary state distribution $\{p_a\}$ over nodes $\{a\}$ that is independent of the initial conditions. The instantaneous probability of the walker being at a node b at time t, $q_b(t)$, is governed by

$$\frac{dq_b(t)}{dt} = \sum_a \omega_{ab} q_a(t) - \sum_a \omega_{ba} q_b(t) = \sum_a \mathbf{\Omega}_{ba} q_a(t) \quad (1)$$

where the time independent rate of transition ω_{ab} from node a to node b is non zero if and only $(a, b) \in E$.

In order to maximize the path entropy, we discretize time into time intervals δt and at a later time, take the limit $\delta t \to 0$. In the discrete time scenario, the transition rates ω_{ab} with units of inverse time are replaced by unitless transition probabilities k_{ab}. The matrix $\mathbf{k}$ of transition probabilities depends on the discretization interval and is given by $\mathbf{k} = e^{\mathbf{\Omega}\delta t} \approx \mathbf{I} + \mathbf{\Omega}\delta t$. The second approximation is accurate only for $\delta t \ll 1$. Here, $\mathbf{I}$ is the identity matrix.

Let us define an ensemble $\{\Gamma\}$ of stationary state paths $\Gamma \equiv \cdots \to a \to b \to c \to d \ldots$ of the walker of total duration T. The entropy of this ensemble is $\mathcal{S}_T = -\sum_\Gamma P(\Gamma) \log P(\Gamma) = -T \sum p_a k_{ab} \log k_{ab}$ (8–10). The time normalized path entropy $\mathcal{S}$ is

$$\mathcal{S} = -\sum_{(a,b)\in E} p_a k_{ab} \log k_{ab}. \quad (2)$$

The sum in Eq. 2 is only taken over edges $(a, b) \in E$ of the network. All summations below are also restricted to edges $(a, b) \in E$ unless otherwise stated.

The constraints on the paths

Any discrete time Markov process with $\{p_a\}$ as the stationary distribution and k_{ab} as transition probabilities satisfies two sets of linear constraints (normalization and stationarity). These constraints are understood as follows: First, from state i at time t, the system *has to* land at one of the states j at time $t + \delta t$. Second, at stationary state, a system in state j at time $t + \delta t$ comes from one of the state i. We have,

$$\sum_b k_{ab} = 1 \ \forall \ a \text{ and } \sum_a p_a k_{ab} = p_b \ \forall \ b. \quad (3)$$

Another important constraint is detailed balance, $p_a k_{ab} = p_b k_{ba}$. Below, we will see how detailed balanced constrained can be applied explicitly (9).

We introduce a node-connectivity variable N such that $N_{aa} = 0$, $N_{ab} = 1$ *if* $(a, b) \in E$. The path ensemble average of N over all trajectories Γ is given by (8, 9)

$$\langle N \rangle = \sum_{(a,b)\in E} p_a k_{ab} N_{ab}. \quad (4)$$

$\langle N \rangle$ is the mean number of transitions in a single time step δt. Below we use $\langle N \rangle$ to take the desired limit $\delta t \to 0$ in order to convert the discrete time Markov chain to a continuous time Markov process.

Also, we constrain the ensemble average $\langle r^i_{ab} \rangle$ of arbitrary dynamical rate variables r^i_{ab}. Examples of dynamical constraints include the average distance travelled by a particle diffusing on an energy landscape per unit time, the number of spin flips per unit time in a spin glass model, the average number of reactions per unit time, etc. We assume $r^i_{aa} = 0$ and $r^i_{ab} = 0$ if $(a, b) \notin E$. $r^i_{aa} = 0$ is a mere convenience and not a strict requirement of our development (see supplementary materials for details). Given that the network is directed, in general we have $N_{ab} \neq N_{ba}$ and $r^i_{ab} \neq r^i_{ba}$.

Maximizing the path entropy, subject to constraints.

We now maximize the path entropy $\mathcal{S}$ (Eq. 2) with respect to transition probabilities k_{ab} subject to constraints imposed by Eq. 3 and Eq. 4. Using the method of Lagrange multipliers, we write the unconstrained Caliber $\mathcal{C}$ (6, 11)

$$\mathcal{C} = \mathcal{S} + \sum_a m_a \left(\sum_b p_a k_{ab} - p_a \right) + \sum_b l_b \left(\sum_a p_a k_{ab} - p_b \right) - \gamma \left(\sum_{a,b} p_a k_{ab} N_{ab} - \langle N \rangle \right) - \sum_i \rho_i \left(\sum_{a,b} p_a k_{ab} r^i_{ab} - \langle r^i_{ab} \rangle \right). \quad (5)$$

We maximize the Caliber with respect to k_{ab} and derive a closed form expression for the transition probabilities (see supplementary materials for a detailed derivation)

$$k_{ab} = \mu \delta t \sqrt{\frac{f_b g_a}{f_a g_b}} \sqrt{\frac{p_b}{p_a}} \mathbf{\Delta}_{ab} \ \textit{if} \ (a, b) \in E \text{ and} \quad (6)$$

and $k_{aa} = 1 - \sum_b k_{ab}$. Here $\mu \delta t = e^{-\gamma}$ and $\mathbf{\Delta}_{ab} = e^{-\sum_i \rho_i r^i_{ab}}$ *if* $(a, b) \in E$ and zero otherwise. Constants f_a and g_a are determined from self-consistent equations

$$f_a = \sum_b \mathbf{\Delta}_{ab} \sqrt{\frac{f_b p_b}{g_b}} \text{ and } g_a = \sum_b \mathbf{\Delta}_{ba} \sqrt{\frac{g_b p_b}{f_b}} \quad (7)$$

Since $\mathbf{k} = \mathbf{I} + \mathbf{\Omega}\delta t$ as $\delta t \to 0$. We take the limit and get

$$\omega_{ab} = \mu\sqrt{\frac{f_b g_a}{f_a g_b}}\sqrt{\frac{p_b}{p_a}}\mathbf{\Delta}_{ab} \; if \; (a,b) \in E \tag{8}$$

and $\omega_{aa} = -\sum_b \omega_{ab}$. Eqs. 8 is the most general result of this work. Since constants f_a and g_a are determined from *global* self-consistent equations Eq. 7, the transition rate ω_{ab} between any two states a and b depends on the structure of the entire network.

Detailed balance

Detailed balance constraint requires $p_a\omega_{ab} = p_b\omega_{ba}$. Thus, if both (a,b) and (b,a) are not in E, $\omega_{ba} = \omega_{ab} = 0$ and the connection between (a,b) can be removed. Consequently, we assume that the network G is undirected i.e. $(a,b) \in E \Rightarrow (b,a) \in E$ and vice versa. Imposing detailed balance constraints on transition probabilities k_{ab} is equivalent to constraining symmetrized forms of the dynamical variables $r^{i\dagger}_{ab} = \frac{1}{2}\left(r^i_{ab} + r^i_{ba}\right)$ (see (9) for a proof). In this case, we have $\mathbf{\Delta} = \mathbf{\Delta}^{\mathrm{T}}$ and $\bar{f} = \bar{g}$. Thus

$$\omega_{ab} = \mu\sqrt{\frac{p_b}{p_a}}\mathbf{\Delta}_{ab} \; if \; (a,b) \in E. \tag{9}$$

It is easy to check that Eq. 9 satisfies detailed balance. Interestingly, the transition rates of a detailed balanced Markov process are determined entirely locally from the properties of states a and b alone and does not depend on the global structure of the network. In contrast, the same transition rate depends on the structure of the entire network when detailed balance is not satisfied (see Eq. 8).

When we constrain only the mean jump rate ($\mathbf{\Delta}_{ab} = 1$ *if* $(a,b) \in E$ and zero otherwise), the transition rates are given by the simple expression:

$$\omega_{ab} = \mu\sqrt{\frac{p_b}{p_a}} \; if \; (a,b) \in E. \tag{10}$$

This result should be contrasted with other popular functional forms of the transition rates that satisfy a prescribed stationary distribution, e.g. Glauber dynamics (12) and Metropolis dynamics (13). For example, historically, Glauber developed his dynamics to study magnetic spins (12). The particular form of the transition rates was motivated by "a desire for simplicity" (12). The derivation of square root dynamics presented here, while not resorting to a *ad hoc* prescription for transition rates, follows the same intuitive principle i.e. finding the simplest model that is consistent with a given set stationary and dynamical constraints and surprisingly predicts a dynamics that is *qualitatively* different from Glauber and Metropolis dynamics.

Application to theories of chemical reaction rates

Our predicted square-root dynamics has an interesting interpretation in theories of chemical reaction rates. In organic chemistry, so-called *extra-thermodynamic relationships* express empirical observations about how the rates and mechanisms of certain types of chemical reactions are related to their equilibria. These go by the name of the Bronsted relation (often applied to acid-base catalysis) (14), the Polanyi relationship for surface catalysis (15), Marcus theory for electron transfer reactions in solution (16), or Φ-value analysis for protein folding (17). In short, these approaches express that rates are related to equilibrium constants in the form of

$$\omega(x) = cK(x)^{\alpha} \tag{11}$$

where x is a variable representing a systematic change, such as a series of different acids of different pK_a's, $K(x)$ is the equilibrium constants for that series, and $\omega(x)$ are the rates of reaction. α is a value that usually ranges from 0 to 1, expressing the degree of resemblance of the transition state to the products of the reaction.

Now consider a two state system where state a are reactants and state b are products such that $p_b > p_a$. We have neglected the population in the transition state since it is much smaller than the state populations at equilibrium. We have equilibrium constant $K = p_b/p_a > 1$. If no dynamical constraints are imposed, the transition rate ω_{ab} of the $a \to b$ reaction according to the MaxCal process is given by $\omega_{ab} = \mu\sqrt{\frac{p_b}{p_a}} = \mu K^{1/2}$. In short, our MaxCal approach predicts that $\alpha = 1/2$ in Eq. 11 is the most parsimonious assumption for the reaction mechanism prior to any knowledge of global rate information, implying that the transition state is halfway between reactants and products.

ILLUSTRATING EQS. 8 AND 9 WITH TWO EXAMPLES.

We test the model predictions on a problem of correlated expression of two genes that are transcribed by the same promoter and on a problem of the equilibrium conformations of a small peptide.

Expression dynamics of two correlated genes

It is difficult to infer the underlying regulatory architecture of large biochemical networks from stationary populations of the components such as mRNAs and proteins. Here, we show how Eq. 8 can accurately predict the full chemical master equation (CME) from stationary distributions and overall rate parameters. Consider a biochemical circuit where two genes are adjacent to a

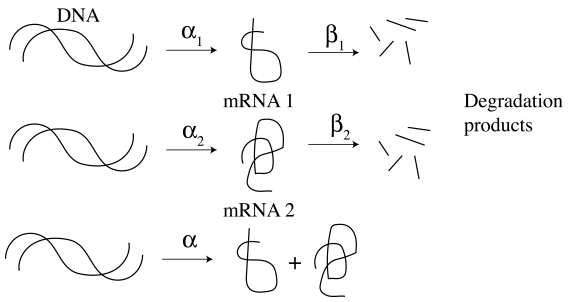

FIG. 1. Schematic of two genes that are simultaneously transcribed by the same promoter. Rates α_1, α_2 ,and α represent the synthesis rates and rates β_1 and β_2 represent the degradation rates.

constitutively expressing promoter region. We assume that the genes are either transcribed individually or simultaneously and that they are degraded individually. In this toy example, two stochastic variables are correlated, namely the copy numbers of the two mRNA molecules. We first construct a CME to mimic the biochemical circuit *in silico* (Fig. 1). There are 5 rate parameters in the CME: 3 synthesis rates α_1, α_2, and α and two degradation rates β_1 and β_2. If n_1 and n_2 are the number of molecules of the first and the second mRNA, the chemical master equation describing the system has the following form

$$\begin{aligned}\frac{dp(n_1,n_2;t)}{dt} &= \alpha_0\left(p(n_1-1,n_2)-p(n_1+1,n_2)\right)\\ &+\alpha_1\left(p(n_1,n_2-1)-p(n_1,n_2+1)\right)\\ &+\alpha\left(p(n_1-1,n_2-1)-p(n_1,n_2)\right)\\ &+\beta_1\left((n_1+1)p(n_1+1,n_2)-n_1p(n_1,n_2)\right)\\ &+\beta_2\left((n_2+1)p(n_1,n_2+1)-n_2p(n_1,n_2)\right)\end{aligned} \tag{12}$$

Here, $p(n_1,n_2;t)$ is the instantaneous probability of having n_1 and n_2 molecules of mRNA 1 and 2 respectively at time t. The terms correspond to individual synthesis of mRNA 1 and mRNA 2, the simultaneous synthesis of both mRNAs, and the degradation of mRNA 1 and mRNA 2.

We assume that we can experimentally estimate the joint probability distribution $p_{ss}(n_1,n_2)$ of the mRNA copy numbers at steady state. Additionally, we also assume that we have two reporters that count the total number of expression events and the total number of degradation events respectively. Note that the reporters are agnostic to which of the two RNAs has been synthesized or degraded. With only these three pieces of information, can we estimate the transition rate matrix of the system?

We choose the following parameters for the CME: $(\alpha_1,\alpha_2,\alpha,\beta_1,\beta_2)=(1,0.5,2.5,5,10)$. The choice ensures that the number of any of the two mRNA molecules is limited to < 6. This results in a small system size in this proof of principle work where the total number of states is $6\times 6 = 36$. In Fig. 2 panels A and B, we show the numerically observed joint distribution $p_{ss}(n_1,n_2)$. The correlated pattern of expression is apparent in panel B; the probability $p_{ss}(n_2|n_1)$ depends on n_1. The higher the value of n_1, the higher n_2 values become more probable as a result of the correlated expression.

In the CME, while there are only 5 rate parameters, there are 145 possible transitions. There are 30 + 30 transitions that correspond to synthesis of mRNA 1 or 2. There are 30 + 30 transitions that correspond to degradation of mRNA 1 or 2 and there are 25 transitions that correspond to simultaneous synthesis of both mRNAs. Each transition has its own rate constant but not all rate constants are independent of each other. For example, the transition $(2,3)\rightarrow(2,2)$ has a rate constant $3\beta_2$ and the transition $(2,4)\rightarrow(2,3)$ has a rate constant $4\beta_2$. Moreover, many other transition rates are equal to each other, for example, the rates for transitions that increase the first mRNA copy number such as $(1,2)\rightarrow(2,2)$ and $(3,1)\rightarrow(4,1)$ are equal to α_1 and so on.

We find that Eq. 8 accurately gives the full 145 rate parameters, without the knowledge of the differential rates of synthesis and degradation of the two mRNAs (see supplementary materials for details of the fitting procedure) (see Fig. 3). Taken to larger scale, it implies that using steady state data on cell-to-cell variability in gene expression and a few overall kinetic measurements, we can a full set of chemical master equations and infer regulatory details. This infered model is optimal in the Maximum Caliber sense.

Dynamics of a small peptide

Second, we study the equilibrium dynamics of metastable states of a small peptide comprising 7 alanine amino-acid residues. A metastable state is an ensemble of geometrically and dynamically proximal microstructures that have a significant net population. Classifying protein structures into their metastable states is an active area of research (5, 18–20).

As a test, we compare to previous extensive MD simulations of this peptide (21), which led to the identification of 32 metastable conformations (20) (see Fig. 4). From that MD simulation, we took the states as defined by Jain and Stock and estimated the relative probability p_a of each metastable state a as the fraction of time points in the full trajectory when the peptide was in that state. We also estimated the transition probabilities k_{ab} as the fraction of the events in the trajectory when the peptide was in state a at time t and transitioned into state b at time $t+\delta t$ ($\delta t = 1$ ps in a MD simulation of total duration $T = 800$ ns).

To predict the transition probabilities k_{ab}, we first need to estimate the transition rate matrix $\mathbf{\Omega}$ using Eq. 9. In

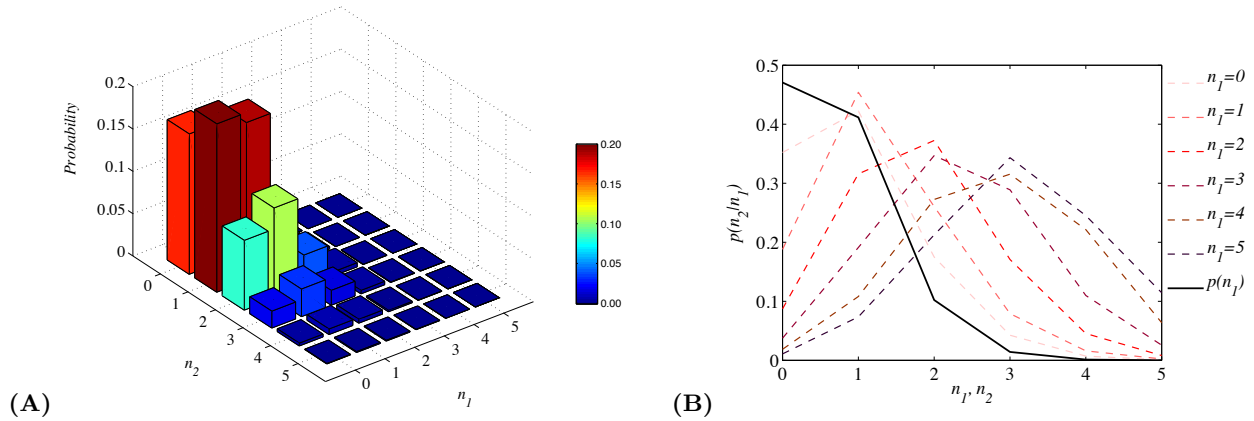

FIG. 2. Panel A: The joint stationary state distribution $p_{ss}(n_1, n_2)$ when the parameters are set at $(\alpha_1, \alpha_2, \alpha, \beta_1, \beta_2) = (1, 0.5, 2.5, 5, 10)$. Panel B: The conditional probability $p(n_2|n_1)$ at different values of n_1.

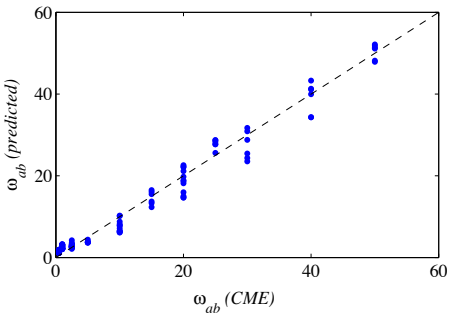

FIG. 3. Predicted rates vs. toy model values for all 145 possible transitions in the gene expression model.

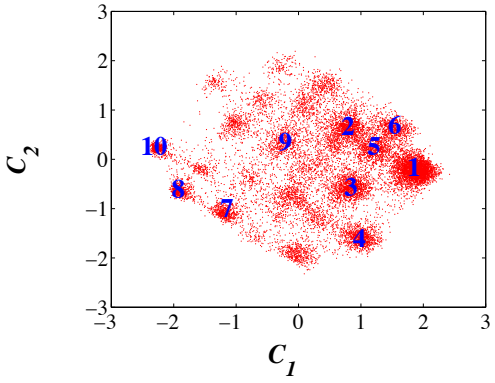

FIG. 4. Two principal coordinates from molecular dynamics simulations of the alanine-7 peptide (21), showing that the microscopic structures can be lumped into well separated metastable states. The 10 most populated states are labelled in decreasing order of state probability.

order to guess the functional form of the transition rate matrix, we need to identify a dynamical constraint. The accuracy of our predictions depends on how well the dynamical constraint captures the diffusion of the peptide on the free energy landscape. Unfortunately, there is no systematic procedure to guess a 'good' constraint variable. This is a common aspect of Maximum-Entropy methods. See (22–24) for a discussion on the role of constraints in maximum entropy methods.

To guess the dynamical constraint, we make two observations: the transition rate decreases when 1) when the average conformational separation between states is increased, keeping the stationary probabilities and the free energy barrier constant and 2) when the free energy barrier is increased, keeping the average conformational distance and the stationary state probabilities constant. As a first guess, we only model the effect of geometric separation and neglect the free energy barrier. We choose a simple geometric constraint r_{ab}: for any two metastable states a and b and microstates x and y such that $x \in a$ and $y \in b$, $r_{ab} = \left(\int_{x\in a, y\in b} ||x-y||_2 dxdy\right)^{\frac{1}{2}}$. r_{ab} is the mean distance between a randomly chosen microstate point x in metastable state a and randomly chosen microstate point y in metastable state b. Similarly, r_{aa} is the average distance between any two microstates x and y within a macrostate a; it represents the 'volume' of macrostate a. The distance between any two microstates x and y is defined as the Euclidean distance between their internal coordinate representation (20). The dynamical average $\langle r_{ab}\rangle$ represents the average distance travelled by the microstates of the peptide per unit time.

Using Eq. 9, the numerically estimated stationary probabilities, and the chosen geometric constraint, we arrive at the functional form of the transition rate matrix $\mathbf{\Omega}$. The transition matrix had two free parameters, μ the time scale and ρ the Lagrange multiplier associated with $\langle r_{ab}\rangle$. In order to estimate the transition probability matrix $\mathbf{k}$ from the predicted matrix of transition rates $\mathbf{\Omega}$, we need the discretization time scale δt, a third parameter. Since μ and δt can be combined together, we only needed to determine 2 parameters, ρ and $\mu\delta t$. Note that the Maximum Caliber approach guesses the parametric

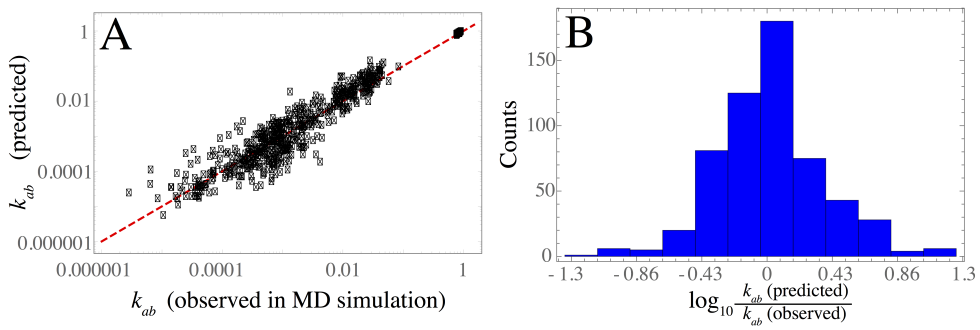

FIG. 5. Panel A: A comparison of transition probabilities $\{k_{ab}\}$ predicted using Eq. 9 and those observed in the MD simulation for the 7 residue peptide shows good agreement over 5 orders of magnitude. Panel B: Histogram of the logarithm of predicted vs observed transition probabilities shows that while the majority of the transition probabilities are predicted very accurately, virtually all transition probabilities are predicted within one order of magnitude.

form of the transition probabilities, one can use any suitable numerical technique and experimental information to estimate the parameters. Since we have access to microscopic data, in this proof of principle study, we used the entire transition probability matrix to fit the two free parameters. Using multiple simulated annealing runs to minimize the total error between known and predicted transition probabilities, we found that the best fits were at $\mu\delta t = 45 \pm 5$ and $\rho = 4.8 \pm 0.6$. Fig. 5 shows that the rates predicted by the Max Cal approach quite accurately capture the correct values obtained from the full MD simulation, over 5 orders of magnitudes of rates.

DISCUSSION

Here, we describe theory that takes a given network, steady-state populations on its nodes, and a couple of global dynamical constraints, and finds the microscopic transition rates among all the nodes. We do this using Maximum Caliber, a Maximum-Entropy-like principle for dynamical systems. A main finding is that the MaxCal transition rates are proportional to the square root of ratios of the state populations. We illustrate our results on a toy gene expression network and peptide conformations, for which we know the correct rates in advance. We believe this treatment could be useful in many areas of network modeling, including in spin-glass models of the immune system (2, 25), colloidal assemblies (3), neuronal networks (1), master-equation models of noisy gene expression (24, 26), and in the browsing behavior of web crawlers on the internet (27).

* Corresponding author. Email: pd2447@columbia.edu

[1] Schneidman, E., M. J. Berry, R. Segev, and W. Bialek, 2006. Weak pairwise correlations imply strongly correlated network states in a neural population. Nature 440:1007–1012.

[2] Shekhar, K., C. F. Ruberman, A. L. Ferguson, J. P. Barton, M. Kardar, and A. K. Chakraborty, 2013. Spin models inferred from patient-derived viral sequence data faithfully describe HIV fitness landscapes. Physical review E 88:062705.

[3] Han, Y., Y. Shokef, A. M. Alsayed, P. Yunker, T. C. Lubensky, and A. G. Yodh, 2008. Geometric frustration in buckled colloidal monolayers. Nature 456:898–903.

[4] Orth, J. D., I. Thiele, and B. Ø. Palsson, 2010. What is flux balance analysis? Nature biotechnology 28:245–248.

[5] Chodera, J. D., and F. Noé, 2014. Markov state models of biomolecular conformational dynamics. Current opinion in structural biology 25:135–144.

[6] Pressé, S., K. Ghosh, J. Lee, and K. A. Dill, 2013. The principles of Maximum Entropy and Maximum Caliber in statistical physics. Rev. Mod. Phys. 85:1115–1141.

[7] Peterson, J., P. D. Dixit, and K. A. Dill, 2013. A maximum entropy framework for nonexponential distributions. Proc. Natl. Acad. Sci. 110:20380–20385.

[8] Filyukov, A., and V. Y. Karpov, 1967. Method of the most probable path of evolution in the theory of stationary irreversible processes. J. Engg. Phys. Thermophys. 13:416–419.

[9] Dixit, P. D., and K. A. Dill, 2014. Inferring microscopic kinetic rates from stationary state distributions. Journal of chemical theory and computation 10:3002–3005.

[10] Cover, T. M., and J. A. Thomas, 2012. Elements of information theory. John Wiley & Sons.

[11] Stock, G., K. Ghosh, and K. A. Dill, 2008. Maximum Caliber: A variational approach applied to two-state dynamics. J. Chem. Phys. 128:194102.

[12] Glauber, R. J., 1963. Time-dependent statistics of the Ising model. Journal of mathematical physics 4:294.

[13] Mariz, A., H. Herrmann, and L. de Arcangelis, 1990. Comparative study of damage spreading in the Ising model using heat-bath, glauber, and metropolis dynamics. Journal of Statistical Physics 59:1043–1050.

[14] Leffler, J. E., and E. Grunwald, 2013. Rates and equilibria of organic reactions: as treated by statistical, thermodynamic and extrathermodynamic methods. Courier Corporation.

[15] Michaelides, A., Z.-P. Liu, C. Zhang, A. Alavi, D. A. King, and P. Hu, 2003. Identification of general linear relationships between activation energies and enthalpy changes for dissociation reactions at surfaces. Journal of the American Chemical Society 125:3704–3705.

[16] Marcus, R. A., 1968. Theoretical relations among rate constants, barriers, and Brønsted slopes of chemical reactions. The Journal of Physical Chemistry 72:891–899.

[17] Matouschek, A., and A. R. Fersht, 1993. Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behavior in the transition state of protein folding. Proceedings of the National Academy of Sciences 90:7814–7818.

[18] Lane, T. J., G. R. Bowman, K. Beauchamp, V. A. Voelz, and V. S. Pande, 2011. Markov state model reveals folding and functional dynamics in ultra-long MD trajectories. Journal of the American Chemical Society 133:18413–18419.

[19] Prinz, J.-H., H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, 2011. Markov models of molecular kinetics: Generation and validation. The Journal of chemical physics 134:174105.

[20] Jain, A., and G. Stock, 2012. Identifying metastable states of folding proteins. Journal of Chemical Theory and Computation 8:3810–3819.

[21] Altis, A., M. Otten, P. H. Nguyen, R. Hegger, and G. Stock, 2008. Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis. The Journal of chemical physics 128:245102.

[22] Caticha, A., and R. Preuss, 2004. Maximum entropy and Bayesian data analysis: Entropic prior distributions. Phys. Rev. E 70:046127.

[23] Caticha, A., 2012. Entropic Inference: some pitfalls and paradoxes we can avoid. arXiv preprint arXiv:1212.6967 .

[24] Dixit, P. D., 2013. Quantifying Extrinsic Noise in Gene Expression Using the Maximum Entropy Framework. Biophys. J. 104:2743–2750.

[25] Mora, T., A. M. Walczak, W. Bialek, and C. G. Callan, 2010. Maximum entropy models for antibody diversity. Proceedings of the National Academy of Sciences 107:5405–5410.

[26] Paulsson, J., 2004. Summing up the noise in gene networks. Nature 427:415–418.

[27] Baldi, P., P. Frasconi, and P. Smyth, 2003. Modeling the Internet and the Web. Probabilistic methods and algorithms .

SUPPLEMENTARY MATERIALS

Deriving the maximum Caliber transition rates

For notational simplicity, we derive Eq. 8 for one dynamical constraints r_{ab}. Differentiating the Caliber in Eq. 5 with respect to k_{ab} and setting the derivative to zero,

$$\begin{aligned} p_a(\log k_{ab} + 1) &= m_a p_a + l_b p_a - \gamma N_{ab} p_a - \rho r_{ab} p_a \\ \Rightarrow k_{ab} &= \frac{\beta_a}{p_a} \lambda_b \mathbf{W}_{ab} \end{aligned} \tag{13}$$

where $\frac{\beta_a}{p_a} = e^{m_a - 1}$, $\lambda_b = e^{l_b}$, and $\mathbf{W}_{ab} = e^{-\gamma N_{ab} - \rho r_{ab}}$.

For a given value of γ and ρ, the Lagrange multipliers βs and λs are determined by self consistently solving Eqs. 3. We have

$$\begin{aligned} \sum_b k_{ab} = 1 \Rightarrow \sum_b \mathbf{W}_{ab} \lambda_b &= \frac{p_a}{\beta_a} \\ \text{and} \quad \sum_a p_a k_{ab} = p_b \Rightarrow \sum_b \mathbf{W}_{ba} \beta_b &= \frac{p_a}{\lambda_a}. \end{aligned} \tag{14}$$

To further simplify Eqs. 14, let usidentify $\bar{\lambda}$ and $\bar{\beta}$ as the vectors of Lagrange multipliers and define $\mathcal{D}[\bar{x}]_a = \frac{p_a}{x_a}$, a non-linear operator on vectors $\bar{x}$. We have

$$\begin{aligned} \mathbf{W}\bar{\lambda} = \mathcal{D}[\bar{\beta}] \text{ and } \mathbf{W}^{\mathrm{T}}\bar{\beta} = \mathcal{D}[\bar{\lambda}] \\ \Rightarrow \mathcal{D}[\mathbf{W}\bar{\lambda}] = \bar{\beta} \text{ and } \mathcal{D}[\mathbf{W}^{\mathrm{T}}\bar{\beta}] = \bar{\lambda} \end{aligned} \tag{15}$$

Now, we write $\mathbf{W} = \mathbf{I} + \mu\delta t \mathbf{\Delta}$ where $e^{-\gamma} = \mu\delta t$, $\mathbf{\Delta}_{ab} = 0$ if $(a,b) \notin E$, and $\mathbf{\Delta}_{ab} = e^{-\rho r_{ab}}$ when $(a,b) \in E$. We have

$$\frac{p_a}{\lambda_a + \mu\delta t f_a} = \beta_a \text{ and } \frac{p_a}{\beta_a + \mu\delta t g_a} = \lambda_a \tag{16}$$

where

$$\bar{f} = \mathbf{\Delta}\bar{\lambda} \text{ and } \bar{g} = \mathbf{\Delta}^{\mathrm{T}}\bar{\beta}. \tag{17}$$

We solve these two equations algebraicly recognizing that f_a and g_a do not directly depend on λ_a and β_a. The positive roots are

$$\lambda_a = \frac{\sqrt{f_a}\sqrt{g_a}\sqrt{\delta t^2 f_a g_a \mu^2 + 4p_a} - \delta t f_a g_a \mu}{2g_a} \tag{18}$$

$$\beta_a = \frac{\sqrt{f_a}\sqrt{g_a}\sqrt{\delta t^2 f_a g_a \mu^2 + 4p_a} - \delta t f_a g_a \mu}{2f_a}. \tag{19}$$

Given that we're only interested in transition probabilities k_{ab} up to orders of δt and $k_{ab} = \mu\delta t \frac{\beta_a}{p_a} \lambda_b \Delta_{ab}$ already has a δt term, we only need to the zeroth order terms for $\bar{\lambda}$ and $\bar{\beta}$ as $\delta t \to 0$. We solve the algebraic equations 16 simultaneously for λ_a and β_a in terms of f_a, g_a, p_a, and dt. We then take only the zeroth order terms in δt (see supplementary materials for details). We have

$$\lambda_a = \sqrt{\frac{f_a p_a}{g_a}} \text{ and } \beta_a = \sqrt{\frac{g_a p_a}{f_a}} \tag{20}$$

Eq. 20 and Eq. 17 can be self-consistently solved for $\bar{\lambda}$ and $\bar{\beta}$ and we can obtain $\bar{f} = \mathbf{\Delta}\bar{\lambda}$ and $\bar{g} = \mathbf{\Delta}^{\mathrm{T}}\bar{\beta}$.

When dynamical constraints are such that $r_{aa} \neq 0$

As mentioned in the main text, we have assumed for convenience that the dynamical constraints r_{ab} is such that $r_{aa} = 0$. Here, we show that when $r_{aa} \neq 0$, the maximum entropy problem is equivalent to constraining a modified constraints $r^{\dagger}_{ab} = r_{ab} - \frac{1}{2}(r_{aa} + r_{bb})$.

We start with recognizing $\mathbf{W} = \mathbf{J} + \mu\delta t\mathbf{\Delta}$ as above where $\mathbf{J}$ is a diagonal matrix such that $\mathbf{J}_{aa} = e^{-\rho r_{aa}}$. We have

$$\frac{p_a}{j_a\lambda_a + \mu\delta t f_a} = \beta_a \text{ and } \frac{p_a}{j_a\beta_a + \mu\delta t g_a} = \lambda_a \quad (21)$$

where $j_a = \mathbf{J}_{aa}$ and $\bar{f} = \mathbf{\Delta}\bar{\lambda}$ and $\bar{g} = \mathbf{\Delta}^{\mathrm{T}}\bar{\beta}$. Again solving to zeroth order in δt, we get

$$\lambda_a = \sqrt{\frac{f_a p_a}{g_a j_a}} \text{ and } \beta_a = \sqrt{\frac{p_a g_a}{f_a j_a}}. \quad (22)$$

Finally, the transition probability k_{ab} is given by

$$\begin{aligned} k_{ab} &= \mu\delta t\frac{\beta_a}{p_a}\lambda_b e^{-\rho r_{ab}} = \mu\delta t\sqrt{\frac{p_b}{p_a}}\sqrt{\frac{f_b g_a}{f_a g_b}}\frac{1}{\sqrt{j_a j_b}}e^{-\rho r_{ab}} \\ &= \mu\delta t\sqrt{\frac{p_b}{p_a}}\sqrt{\frac{f_b g_a}{f_a g_b}}e^{-\left(\rho r_{ab} - \frac{1}{2}(r_{aa}+r_{bb})\right)} \\ &= \mu\delta t\sqrt{\frac{p_b}{p_a}}\sqrt{\frac{f_b g_a}{f_a g_b}}e^{-\rho r^{\dagger}_{ab}} \end{aligned} \quad (23)$$

where $r^{\dagger}_{ab} = r_{ab} - \frac{1}{2}(r_{aa} + r_{bb})$. Comparing Eq. 23 to Eq. 8, it is clear that the problem of constraining a dynamical constraint r_{ab} such that r_{aa} is non-zero is equivalent to constraining a modified dynamical constraint $r^{\dagger}_{ab} = r_{ab} - \frac{1}{2}(r_{aa} + r_{bb})$. Note that by definition, $r^{\dagger}_{aa} = 0$. Thus, for convenience, we assume that this transformation is already performed and $r_{aa} = 0$.

Constructing and fitting the maximum entropy Markov proces for the gene network

From the numerical experiments, we obtain accurate estimate of the stationary state probability distribution $p_{ss}(n_1, n_2)$. Let $a \equiv (n_1, n_2)$ and $b \equiv (n_1^{\dagger}, n_2^{\dagger})$ be any two states of the system. We know that there is a directed edge from state a to state b *iff*

1. Synthesis or degradation of mRNA 1: $n_1 = n_1^{\dagger} \pm 1$ and $n_2 = n_2$
2. Synthesis or degradation of mRNA 2: $n_1 = n_1^{\dagger}$ and $n_2 = n_2^{\dagger} \pm 1$
3. Simultaneous synthesis: $n_1^{\dagger} = n_1{+}1$ and $n_2^{\dagger} = n_2{+}1$

As mentioned in the main text, the 'experiments' tell us the total number of degradation and synthesis events per unit time but do not have the ability to distinguish between the two mRNAs. We constrain two quantities, the number of degradation events per unit time and the number of synthesis events per unit time. Accordingly, we set $\mathbf{\Delta}$ in Eq. 8 as follows

1. No edge between nodes a and b: $\mathbf{\Delta}_{ab} = 0$ if $(a, b) \notin E$
2. Synthesis events: $\mathbf{\Delta}_{ab} = \eta$ if $n_1 = n_1^{\dagger}$ and $n_2 + 1 = n_2^{\dagger}$ or $n_1 + 1 = n_1^{\dagger}$ and $n_2 = n_2^{\dagger}$ or $n_1 + 1 = n_1^{\dagger}$ and $n_2 + 1 = n_2^{\dagger}$
3. Degradation of mRNA 1: $\mathbf{\Delta}_{ab} = n_1\zeta$ if $n_1 = n_1^{\dagger} + 1$ and $n_2 = n_2^{\dagger}$
4. Degradation of mRNA2: $\mathbf{\Delta}_{ab} = n_2\zeta$ if $n_1 = n_1^{\dagger}$ and $n_2 = n_2^{\dagger} + 1$

Here, $\eta \geq 0$ and $\zeta \geq 0$ are exponentials of Lagrange multipliers similar to those used in the main text. The Lagrange rate constant μ in Eq. 8 is assumed to be absorbed in the Lagrange multipliers. The next step is to determine $\bar{f}$, $\bar{g}$, $\bar{\lambda}$, and $\bar{\beta}$ using numerically estimated $p_{ss}(n_1, n_2)$ and $\mathbf{\Delta}$. In order to determine the transition rate matrix $\mathbf{\Omega}$ for any value of η and ζ, we solve Eq. 17 and Eq. 20 self consistently.

Given that we have access to all rate constants in this proof of principles work, we minimize the error between the known rate constants of Eq. 12 and those predicted by Eq. 8 by varying η and ζ using a simulated annealing protocol. We find that $\eta \approx 0.4$ and $\zeta \approx 26.5$ resulted in the best agreement between the known and the predicted rate constants. Multiple simulated annealing runs predicted rates that were identical to the ones shown in the main text. Note that although we used the entire transition matrix to learn the Lagrange multipliers, it is equally possible to learn them from global dynamical constraints.